View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Journal of Number Theory 132 (2012) 324–331

Contents lists available at SciVerse ScienceDirect

Journal of Number Theory

www.elsevier.com/locate/jnt

Digit sums of binomial sums

Arnold Knopfmacher [a], Florian Luca [b,a,*]

[a] The John Knopfmacher Centre for Applicable Analysis and Number Theory, University of the Witwatersrand, P.O. Wits 2050, South Africa
[b] Instituto de Matemáticas, Universidad Nacional Autonoma de México, C.P. 58089, Morelia, Michoacán, Mexico

ARTICLE INFO

Article history:
Received 30 March 2011
Accepted 11 July 2011
Available online 4 October 2011
Communicated by Michael A. Bennett

Keywords:
Sum of digits
Binomial coefficients
Linear forms in logarithms

ABSTRACT

Let $b \geqslant 2$ be a fixed positive integer and let $S(n)$ be a certain type of binomial sum. In this paper, we show that for most n the sum of the digits of $S(n)$ in base b is at least $c_0 \log n/(\log\log n)$, where c_0 is some positive constant depending on b and on the sequence of binomial sums. Our results include middle binomial coefficients $\binom{2n}{n}$ and Apéry numbers A_n. The proof uses a result of McIntosh regarding the asymptotic expansions of such binomial sums as well as Baker's theorem on lower bounds for nonzero linear forms in logarithms of algebraic numbers.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Let $\mathbf{r} := (r_0, r_1, \ldots, r_m)$ be fixed nonnegative integers and put

$$S(n) := \sum_{k=0}^{n} \binom{n}{k}^{r_0} \binom{n+k}{k}^{r_1} \cdots \binom{n+km}{k}^{r_m} \quad \text{for } n = 0, 1, \ldots. \tag{1}$$

In what follows, we put $r := r_0 + \cdots + r_m$. We assume that $r_0 > 0$. When $\mathbf{r} = (1)$, we simply get that

$$S(n) = \sum_{k=0}^{n} \binom{n}{k} = 2^n \quad \text{for all } n \geqslant 0. \tag{2}$$

* Corresponding author at: Instituto de Matemáticas, Universidad Nacional Autonoma de México, C.P. 58089, Morelia, Michoacán, Mexico.

E-mail addresses: Arnold.Knopfmacher@wits.ac.za (A. Knopfmacher), fluca@matmor.unam.mx (F. Luca).

0022-314X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jnt.2011.07.004

When $\mathbf{r} = (2)$, we get that

$$S(n) = \sum_{k=0}^{n} \binom{n}{k}^2 = b_n \quad \text{for all } n \geqslant 0, \tag{3}$$

where $b_n = \binom{2n}{n}$ is the middle binomial coefficient, while when $\mathbf{r} = (2, 2)$, we get that

$$S(n) = \sum_{k=0}^{n} \binom{n}{k}^2 \binom{n+k}{k}^2 = A_n \quad \text{for all } n \geqslant 0, \tag{4}$$

where A_n is the nth Apéry number.

Let $b \geqslant 2$ be any integer and put $s_b(m)$ for the sum of the base b digits of the positive integer m. Lower bounds for $s_b(m)$ when m runs through the members of a sequence with some interesting combinatorial meaning have been investigated before. For example, it follows from a result of Stewart ([8]; see also [1] for a slightly more general result), that the inequality

$$s_b(F_n) > c_1 \frac{\log n}{\log\log n} \tag{5}$$

holds for all $n \geqslant 3$ with some positive constant c_1 depending on b, where F_n is the nth Fibonacci number given by $F_0 := 0$, $F_1 := 1$ and $F_{n+2} := F_{n+1} + F_n$ for all $n \geqslant 0$. In [2], it is shown that the inequality

$$s_b(n!) > c_2 \log n \tag{6}$$

holds for all $n \geqslant 1$, where c_2 is some positive constant depending on b. In [4], it is shown that the inequality

$$s_b(b_n) \geqslant \varepsilon(n)\sqrt{\log n} \tag{7}$$

holds on a set of n of asymptotic density 1, where $\varepsilon(n)$ is any function tending to zero when n tends to infinity. A similar result holds with b_n replaced by $c_n := b_n/(n+1)$ for $n = 0, 1, \ldots,$ which is the nth Catalan number.

In [3], it is shown that there is some positive constant c_3 depending on b such that the inequality

$$s_b(A_n) > c_3 \left(\frac{\log n}{\log\log n}\right)^{1/4} \tag{8}$$

holds on a set of n of asymptotic density 1. The proofs of such results use a variety of methods from number theory, such as elementary methods, sieve methods, lower bounds for nonzero linear forms in logarithms of algebraic numbers and the subspace theorem of Evertse, Schlickewei and Schmidt. Here, we add on the literature on the topic and prove the following theorem.

Theorem 1. *For any $\mathbf{r} \neq (1)$, there exists a positive constant $c_0 := c_0(b, \mathbf{r})$ depending on both b and $\mathbf{r}$ such that the inequality*

$$s_b\bigl(S(n)\bigr) > c_0 \frac{\log n}{\log\log n}$$

holds on a set of positive integers n of asymptotic density 1.

When $\mathbf{r} = (1)$, then $S(n) = 2^n$, so either b is a power of 2, in which case $s_b(2^n) = O(1)$, or b is not a power of 2, in which case it follows from Stewart's result [8] that the inequality $s_b(2^n) > c_4 \log n / \log\log n$ holds for all sufficiently large positive integers n with some positive constant c_4 depending on b.

It is easy to see that particular cases of Theorem 4 such as (3) and (4) are improvements upon estimates (7) and (8). While strictly speaking the Catalan numbers c_n are not of the form $S(n)$ for any particular choice of $\mathbf{r}$, the conclusion of Theorem 4 applies to them also as we shall indicate at the proof of Theorem 4. Features of the proof are a complete asymptotic expansions for $S(n)$ due to McIntosh [7] and a lower bound for a nonzero linear form in logarithms of algebraic numbers due to Matveev [6].

Until now, and in what follows, we use $c_0, c_1, \ldots$ for computable positive constants that appear increasingly throughout the paper and which might be absolute or depend on the number b and the vector $\mathbf{r}$. We use the Landau symbol O and the Vinogradov symbols $\ll$, $\gg$ and $\asymp$ with their usual meanings. Recall that $A = O(B)$, $A \ll B$ and $B \gg A$ are all equivalent to the fact that the inequality $|A| \leqslant cB$ holds with some constant c. The constants implied by these symbols in our arguments might depend on the number b and the vector $\mathbf{r}$. Furthermore, $A \asymp B$ means that both $A \ll B$ and $B \ll A$ hold.

2. Preliminary results

We start with McIntosh's asymptotic formula for $S(n)$ (see [7]).

Lemma 2. *For each nonnegative integer p,*

$$S(n) = \frac{\mu^{n+1/2}}{\sqrt{\nu(2\pi\lambda n)^{r-1}}}\left(1 + \sum_{k=1}^{p} \frac{R_k}{n^k} + O\left(\frac{1}{n^{p+1}}\right)\right), \tag{9}$$

where $0 < \lambda < 1$ is defined by

$$1 = \prod_{j=0}^{m} \left(\frac{(1+j\lambda)^j}{\lambda(1+(j-1)\lambda)^{j-1}}\right)^{r_j},$$

$$\mu = \prod_{j=0}^{m} \left(\frac{1+j\lambda}{1+(j-1)\lambda}\right)^{r_j},$$

$$\nu = \sum_{j=0}^{m} \frac{r_j}{(1+(j-1)\lambda)(1+j\lambda)},$$

and each R_k is a rational function of the exponents $r_0, r_1, \ldots, r_m$ and λ.

We shall also need a result of Matveev [6] from transcendental number theory. But first, some notation. For an algebraic number η having

$$F(X) := a_0 \prod_{i=1}^{d} \left(X - \eta^{(i)}\right) \in \mathbb{Z}[X]$$

as minimal polynomial over the integers, the logarithmic height of η is defined as

$$h(\eta) := \frac{1}{d}\left(\log|a_0| + \sum_{i=1}^{d} \log\max\{|\eta^{(i)}|, 1\}\right).$$

With this notation, Matveev [6] proved a deep theorem a particular case of which is the following.

Lemma 3. *Let $\mathbb{K}$ be a real field of degree D, $\eta_1 > 0$, $\eta_2 > 0$ be elements of $\mathbb{K}$, and b_1, b_2 be nonzero integers. Put $B := \max\{|b_1|, |b_2|\}$ and*

$$\Lambda := b_1 \log \eta_1 - b_2 \log \eta_2.$$

Let A_1, A_2 be real numbers such that

$$A_j \geqslant \max\{Dh(\eta_j), |\log \eta_j|, 0.16\}, \quad \textit{for } j = 1, 2.$$

Then there exists an absolute constant c_5 such that if $\Lambda \neq 0$, then

$$\log|\Lambda| > -c_5 D^2(1 + \log D)(1 + \log B)A_1 A_2.$$

3. The proof of Theorem 4

We let x be a large positive real number. We let $\delta > 0$ be sufficiently small to be determined later, and let

$$\mathcal{N}_\delta(x) := \left\{n \in [x, 2x)\colon\ s_b\big(S(n)\big) < \delta \frac{\log x}{\log\log x}\right\}. \tag{10}$$

We need to show that if δ is sufficiently small, then $\#\mathcal{N}_\delta(x) = o(x)$ as $x \to \infty$, for after this the conclusion of Theorem 4 will follow by replacing x by $x/2$, then by $x/4$, and so on, and summing up the resulting estimates.

For $n \in \mathcal{N}_\delta(x)$, we write

$$S(n) = d_1 b^{n_1} + d_2 b^{n_2} + \cdots + d_s b^{n_s}, \tag{11}$$

where $d_1, \ldots, d_s \in \{1, \ldots, b-1\}$ and $n_1 > n_2 > \cdots > n_s$. We let K be some large number depending on b and $\mathbf{r}$ to be determined later, and we put $t := t(n)$ for the smallest index $i \in \{1, 2, \ldots, s-1\}$ such that $b^{n_i - n_{i+1}} > n^K$ if it exists and set $t := s$ otherwise. From the definition of $t(n)$, we see immediately that

$$S(n) = \big(d_1 b^{n_1} + \cdots + d_t b^{n_t}\big)\left(1 + O\left(\frac{1}{n^K}\right)\right) := b^{m(n)} D(n)\left(1 + O\left(\frac{1}{x^K}\right)\right), \tag{12}$$

where $m = m(n) := n_t$ and $D(n) := d_1 b^{n_1 - n_t} + d_2 b^{n_2 - n_t} + \cdots + d_t$.

Let $\mathcal{D}_\delta(x)$ be the subset of all possible values for $D(n)$. Let us find an upper bound for the cardinality of this set. Observe first that

$$t \leqslant s \leqslant s_b\big(S(n)\big) \leqslant \delta \frac{\log x}{\log\log x}.$$

Next, observe that for a fixed t, the vector $(d_1, \dots, d_t)$ can be chosen in $(b-1)^t < x^{(\delta \log b)/\log\log x}$ ways. Finally, we also have that

$$n_{i-1} - n_i \leqslant \frac{K \log n}{\log b} \leqslant \frac{K \log(2x)}{\log b} \quad \text{for } i = 2, \dots, t.$$

Thus, the vector of numbers $(n_1 - n_2, \dots, n_{t-1} - n_t)$ can be chosen in at most

$$\left(\frac{K \log(2x)}{\log b}\right)^{t-1} < \exp\left(\delta\left(\frac{\log x}{\log\log x}\right)\log\left(\frac{K \log(2x)}{\log b}\right)\right) < x^{2\delta}$$

ways, where the last inequality holds for all sufficiently large x. Moreover, the vector of neighboring differences $(n_1 - n_2, n_2 - n_2, \dots, n_{t-1} - n_t)$ determines uniquely the vector of exponents $(n_1 - n_t, n_2 - n_t, \dots, n_{t-1} - n_t)$ appearing in the base b representation of D via the fact that

$$n_i - n_t = \sum_{j=i}^{t-1} (n_j - n_{j+1}) \quad \text{for } i = 1, \dots, t-1.$$

Thus, we get easily that

$$\#\mathcal{D}_\delta(x) \leqslant \left(\frac{\delta \log x}{\log\log x}\right) \times x^{\delta \log b/(\log\log x)} \times x^{2\delta} < x^{3\delta}, \tag{13}$$

where the last inequality holds provided that x is sufficiently large. We now compare relations (12), (9) and get that

$$\mu^n f(n) = b^{m(n)} D(n)\left(1 + O\left(\frac{1}{x^K}\right)\right), \tag{14}$$

where we put

$$f(n) := \sqrt{\frac{\mu}{\nu(2\pi\lambda n)^{r-1}}}\left(1 + \sum_{k=0}^{K-1} \frac{R_k}{n^k}\right). \tag{15}$$

Taking logarithms in Eq. (14), we get that

$$n \log \mu + \log f(n) - m(n) \log b - \log D(n) = O\left(\frac{1}{x^K}\right).$$

We now write

$$\mathcal{N}_\delta(x) = \bigcup_{D \in \mathcal{D}_\delta(x)} \mathcal{N}_{\delta,D}(x), \tag{16}$$

where

$$N_{\delta,D}(x) := \left\{n \in \mathcal{N}_\delta(x):\ D(n) = D\right\}.$$

Assume that $\mathcal{N}_{\delta,D}(x)$ has $T := T_{\delta,D}(x)$ elements and let them be $n_1 < n_2 < \cdots < n_T$. Let $n = n_i$ for some $i \leqslant T-1$ and write $n_{i+1} =: n+k$. Then taking the difference of the relations (16) in n and $n+k$, we get

$$k\log\mu + \big(\log f(n+k) - \log f(n)\big) - \big(m(n+k) - m(n)\big)\log b = O\left(\frac{1}{x^K}\right). \tag{17}$$

If $r = 1$, then $\mathbf{r} = (1)$, a case which is excluded. Thus, $r > 1$ and, in particular, $f(n)$ is not constant. By elementary calculus, we get that

$$\big|f(n+k) - f(n)\big| = k\left|\left[\frac{d}{dz}f(z)\right]_{z=\zeta\in[n,n+k]}\right| \asymp \frac{k}{x^{(r+1)/2}}, \tag{18}$$

where ζ is some point in $[n, n+k]$ the existence of which is guaranteed by the Intermediary Value Theorem. Observe also that $\mu > 1$ because $S(n)$ tends to infinity with n. Put $m := m(n+k) - m(n)$ and

$$\Lambda := k\log\mu - m\log b. \tag{19}$$

We take $K := \lfloor (r+1)/2\rfloor + 1$. If $\Lambda = 0$, we then get by estimates (17) and (18), that

$$\frac{k}{x^{(r+1)/2}} \ll \frac{1}{x^K},$$

which is impossible for $k \geqslant 1$ for large x because $K > (r+1)/2$. Thus, $\Lambda \neq 0$. Then estimates (17) and (18) again show that

$$|\Lambda| \ll \frac{k}{x^{(r+1)/2}}. \tag{20}$$

Since $r \geqslant 2$ and $k \leqslant x$, it follows that the right-hand side above is $o(1)$ as $x \to \infty$. Thus, $|m(n+k) - m(n)| \sim c_6 k \asymp k$ as $x \to \infty$, where $c_6 := (\log\mu)/(\log b)$. Now Matveev's result Lemma 3 applied to Λ, with the choices of parameters $\eta_1 := \mu$, $\eta_2 := b$, $b_1 := k$, $b_2 := m(n+k) - m(n)$, together with the fact that $B = \max\{k, |m(n+k) - m(n)|\} \asymp k$, shows that there exists a positive constant c_7 depending on b and $\mathbf{r}$ (actually, it depends on b and the height of the algebraic number μ, but this last parameter depends on $\mathbf{r}$), such that the inequality

$$|\Lambda| > \frac{1}{k^{c_7}} \tag{21}$$

holds for all sufficiently large x. We may assume that $c_7 > (r-1)/2$. Putting together (20) and (21), we get that

$$\frac{1}{k^{c_7}} \ll \frac{k}{x^{(r+1)/2}},$$

giving $k \gg x^{c_8}$, where $c_8 := (r+1)/(2(1+c_7)) \in (0,1)$. Since the distance between any two elements of $\mathcal{N}_{\delta,D}(x)$ is $\gg x^{c_8}$, and this set is contained in $[x, 2x)$, it follows that

$$\#\mathcal{N}_{\delta,D}(x) \ll x^{c_9},$$

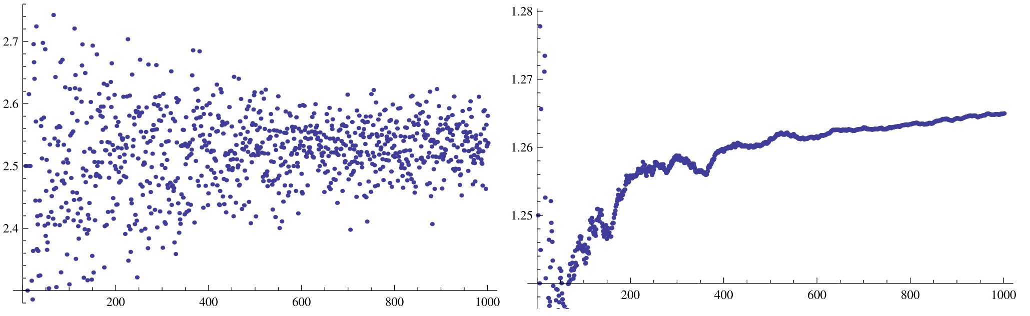

Fig. 1. The binary sum of digits for the Apéry numbers.

where we put $c_9 := 1 - c_8$. This was for an arbitrary $D \in \mathcal{D}_\delta(x)$. Thus, by estimate (13),

$$\#\mathcal{N}_\delta(x) = \sum_{D \in \mathcal{D}_\delta(x)} \#\mathcal{N}_{\delta,D}(x) \leqslant x^{c_9} \#\mathcal{D}_\delta(x) < x^{3\delta + c_9},$$

provided that x is large enough. The exponent of x in the right-most bound above is < 1 provided that $\delta < (1 - c_9)/3 = c_8/3$. So, we can take $c_0 = \delta := c_8/4$ and complete the proof of this theorem.

4. Comments and conjectures

What was important for our argument was not the actual formula for $S(n)$ but the fact that it is an integer which has an asymptotic expansion given as in (9) with some algebraic number $\mu > 1$ and some positive integer $r > 1$. In particular, it also applies to Catalan numbers c_n. We give no further details. The binary expansions of Catalan numbers were also studied in [5].

A problem of further interest would be to find the true rate of growth of $s_b(S(n))$ for large n. In Fig. 1 we show for the case $b = 2$ and $S(n)$ equal to the Apéry numbers, a plot of $s_b(S(n))/n$ for $n = 1, \dots, 1000$ and on the right, a plot of the average digit sum $\frac{1}{n}\sum_{i=1}^{n} s_b(S(i))$ for $n = 1, \dots, 1000$.

Based on these and similar computations in other bases b, we are lead to the following conjectures:

Conjecture 4. *For any* $\mathbf{r} \neq (1)$*, there exists a positive constant* $c_0 := c_0(b, \mathbf{r})$ *depending on both* b *and* $\mathbf{r}$ *such that the inequality*

$$s_b\big(S(n)\big) > c_0 n$$

holds for all positive integers n*.*

Conjecture 5. *For any* $\mathbf{r} \neq (1)$*, there exists a positive constant* $c_1 := c_1(b, \mathbf{r})$ *depending on both* b *and* $\mathbf{r}$ *such that the limit*

$$\frac{1}{n}\sum_{i=1}^{n} s_b\big(S(i)\big) \to c_1$$

holds as $n \to \infty$*.*

Furthermore, we believe that $c_1 = (\frac{\log \mu}{\log b})(\frac{b-1}{4})$. Conjecture 5 with this value of c_1 would follow assuming that the digits of $S(n)$ in base b are uniformly distributed. From our calculations (see Fig. 1), it seems that Conjecture 5 might hold for the Apéry numbers in base 2 with $c_1 = 1.271\dots$, which is in agreement with our prediction since for the Apéry numbers we have $\mu = (1+\sqrt{2})^4$, so $(\frac{\log\mu}{\log 2})(\frac{1}{4}) = \frac{\log(1+\sqrt{2})}{\log 2} = 1.271\ldots$.

Acknowledgments

This paper was written while F.L. was in sabbatical from the Mathematical Institute UNAM from January 1 to June 30, 2011 and supported by a PASPA fellowship from DGAPA.

References

[1] F. Luca, Distinct digits in base b expansions of linear recurrence sequences, Quaest. Math. 23 (2000) 389–404.
[2] F. Luca, The number of non-zero digits of $n!$, Canad. Math. Bull. 45 (2002) 115–118.
[3] F. Luca, I.E. Shparlinski, On the g-ary expansions of Apéry, Motzkin and Schröder numbers, Ann. Comb. 14 (2010) 507–524.
[4] F. Luca, I.E. Shparlinski, On the g-ary expansions of middle binomial coefficients and Catalan numbers, Rocky Mountain J. Math., in press.
[5] F. Luca, P.T. Young, On the binary expansion of the odd Catalan numbers, in: F. Luca, P. Stănică (Eds.), Proceedings of the XIVth International Conference on Fibonacci Numbers and Their Applications, Morelia, Mexico, July 5–9, 2010, in press.
[6] E.M. Matveev, An explicit lower bound for a homogeneous rational linear form in logarithms of algebraic numbers, II, Izv. Ross. Akad. Nauk Ser. Mat. 64 (2000) 125–180 (in Russian); translation in: Izv. Math. 64 (2000) 1217–1269.
[7] R. McIntosh, An asymptotic formula for binomial sums, J. Number Theory 58 (1996) 158–172.
[8] C.L. Stewart, On the representation of an integer in two different bases, J. Reine Angew. Math. 319 (1980) 63–72.